THE BULLETIN OF SYMBOLIC LOGIC
Volume 26, Number 3-4, September-December 2020

GÖDEL'S SECOND INCOMPLETENESS THEOREM: HOW IT IS DERIVED AND WHAT IT DELIVERS

SAEED SALEHI

Abstract. The proofs of Gödel (1931), Rosser (1936), Kleene (first 1936 and second 1950), Chaitin (1970), and Boolos (1989) for the first incompleteness theorem are compared with each other, especially from the viewpoint of the second incompleteness theorem. It is shown that Gödel's (first incompleteness theorem) and Kleene's first theorems are equivalent with the second incompleteness theorem, Rosser's and Kleene's second theorems do deliver the second incompleteness theorem, and Boolos' theorem is derived from the second incompleteness theorem in the standard way. It is also shown that none of Rosser's, Kleene's second, or Boolos' theorems is equivalent with the second incompleteness theorem, and Chaitin's incompleteness theorem neither delivers nor is derived from the second incompleteness theorem. We compare (the strength of) these six proofs with one another.

§1. Introduction and preliminaries. The first incompleteness theorem states the existence of a Π_1-sentence $\boldsymbol{\psi}$ for a given sufficiently strong and recursively enumerable (RE) arithmetical theory T such that if T is consistent, then $\mathbb{N} \vDash \boldsymbol{\psi}$ and $T \nvdash \boldsymbol{\psi}$. If T is, moreover, Σ_1-sound (i.e., every T-provable Σ_1-sentence is true in the standard model of natural numbers $\mathbb{N}$), then we also have $T \nvdash \neg\boldsymbol{\psi}$ (since if we had $T \vdash \neg\boldsymbol{\psi}$, then we would have $\mathbb{N} \nvDash \boldsymbol{\psi}$ by the Σ_1-soundness of T and the fact that $\neg\boldsymbol{\psi}$ is a Σ_1-sentence). The Π_1-sentence $\boldsymbol{\psi}$ depends on the theory T, or more precisely, on how T is presented. An RE theory T may be presented (given) by, for example, an input-free Turing machine (or a program) that outputs a set of axioms for the theory T (after running). It is known that a theory T is RE if and only if it can be defined by a Σ_1-formula; that is, for some Σ_1-formula $\boldsymbol{\xi}(x)$, the set $\{\theta \mid \mathbb{N} \vDash \boldsymbol{\xi}(\ulcorner\theta\urcorner)\}$ axiomatizes T, where θ ranges over the sentences and $\ulcorner\theta\urcorner$ denotes the Gödel code of θ (see, e.g., [13, Theorem 3.3]). By Craig's trick [8], every such theory can be axiomatized by a Δ_0-definable set of axioms (see, e.g., [24, Lemma 2.4] or [29, Section 4.3.4]): if $\boldsymbol{\xi}(x) = \exists y\, \boldsymbol{\zeta}(y,x)$ for some Δ_0-formula $\boldsymbol{\zeta}$, then the Δ_0-formula $\boldsymbol{\tau}(x) = \exists u,v \leqslant x\big[x = (u \wedge [\bar{v} = \bar{v}]) \wedge \boldsymbol{\zeta}(v,u)\big]$ defines another axiomatization for the theory T.

Received January 21, 2019.
2020 *Mathematics Subject Classification*. 03F40, 03F30, 68Q30, 03D32.
Key words and phrases. first incompleteness theorem, second incompleteness theorem, Gödel's proof, Rosser's proof, Kleene's proof, Chaitin's proof, Boolos' proof.

1079-8986/20/2603-4-0005
DOI :10.1017/bsl.2020.22

Any given Δ_0-formula $\boldsymbol{\tau}(x)$, with the only one free variable x, defines the theory $\mathcal{T}\!\mathcal{h}_{\boldsymbol{\tau}} = \{\theta \mid \mathbb{N} \vDash \boldsymbol{\tau}(\ulcorner\theta\urcorner)\}$, where θ ranges over sentences. By some fixed Gödel coding, one can construct a Δ_0-formula $\mathtt{prf}_{\boldsymbol{\tau}}(y,x)$, called *the proof predicate* of $\boldsymbol{\tau}$, in the language of arithmetic which states that "y is (the Gödel code of) a proof in $\mathcal{T}\!\mathcal{h}_{\boldsymbol{\tau}}$ of the sentence (with the Gödel code) x" (see [1, p. 215] or [3, p. 204]). Then *the provability predicate* of a system $\boldsymbol{\tau}$ is the Σ_1-formula $\mathtt{Pr}_{\boldsymbol{\tau}}(x) = \exists y\, \mathtt{prf}_{\boldsymbol{\tau}}(y,x)$, and $\mathtt{Con}_{\boldsymbol{\tau}} = \neg\mathtt{Pr}_{\boldsymbol{\tau}}(\ulcorner 0\neq 0\urcorner)$ is *the consistency statement* of $\boldsymbol{\tau}$. Let us note that a theory may have different axiomatizations, and even one single axiomatization of it may have different defining formulas, and so different proof (and provability) predicates, and different consistency statements.

Let us fix a *Base Theory* $\mathfrak{B}$, which is an RE theory such that:

- The theory $\mathfrak{B}$ is a *sound* extension of Robinson's arithmetic (therefore, the theory $\mathfrak{B}$ is Σ_1-complete, i.e., can prove all the true Σ_1-sentences, and can *strongly represent* all the recursive functions; see [28]).
- The theory $\mathfrak{B}$ can prove the *Derivability* (or *Provability*) *Conditions* of Gödel, Hilbert, Bernays, and Löb (see p. 3 below, and cf. [6, 27]).

Of course, Peano's Arithmetic could be taken for $\mathfrak{B}$, though it is too strong for that. However, Robinson's arithmetic (Q or R) seems too weak to be such a base theory (though, we have no concrete proof for, e.g., Q's weakness at hand). By [1, Proposition 16] the Elementary Arithmetic EA may suffice for us (cf. [29, Remark 6.7] where it is argued that one needs $\mathsf{EA}+\mathsf{B}\Sigma_1$, or equivalently $\mathrm{I}\Delta_1$ by [26], for handling the Σ_1-formulas $\mathtt{Pr}_{\boldsymbol{\tau}}$). To stay on the safe side one can take for $\mathfrak{B}$ the finitely axiomatizable theory $\mathrm{I}\Sigma_1$ (the fragment of Peano's arithmetic where the induction axiom scheme is restricted to Σ_1-formulas). One good reason (other than the ability of $\mathrm{I}\Sigma_1$ to arithmetize the syntax and prove the basic propositions of it, see [11]) is that we will need a variant of the proof predicate, denoted $\overline{\underline{\mathtt{prf}}}_{\boldsymbol{\tau}}(y,x)$, whose all Rosserian sentences are equivalent with each other (in the base theory); and for that Primitive Recursive Arithmetic PRA suffices (see [27, Chapter 6, Theorem 3.6]).

By *a system*, we mean a Δ_0-formula $\boldsymbol{\tau}(x)$, with the only one free variable x, such that $\mathcal{T}\!\mathcal{h}_{\boldsymbol{\tau}} \vdash \mathfrak{B}$. A system is consistent (or Σ_1-sound) when $\mathcal{T}\!\mathcal{h}_{\boldsymbol{\tau}}$ is a consistent (or Σ_1-sound) theory.

A mapping $\mathcal{F}\colon \boldsymbol{\tau} \mapsto \mathcal{F}_{\boldsymbol{\tau}}$ which assigns a Π_1-sentence $\mathcal{F}_{\boldsymbol{\tau}}$ to any given system $\boldsymbol{\tau}$ is called a Π_1-*incompleteness witness* when for every consistent system $\boldsymbol{\tau}$ we have $\mathbb{N} \vDash \mathcal{F}_{\boldsymbol{\tau}}$ and $\mathcal{T}\!\mathcal{h}_{\boldsymbol{\tau}} \nvdash \mathcal{F}_{\boldsymbol{\tau}}$. In this paper, we investigate the Π_1-incompleteness witnesses of Gödel [9], Rosser [22], Kleene (first [15] and second [16]), Chaitin [7] and Boolos [5]. Our purpose is comparing those Π_1-incompleteness witnesses with each other, and with *Gödel's second incompleteness theorem*, which is the following statement:

$$\text{If } \boldsymbol{\tau} \text{ is a consistent system, then } \mathcal{T}\!\mathcal{h}_{\boldsymbol{\tau}} \nvdash \mathtt{Con}_{\boldsymbol{\tau}}. \qquad (\mathbb{G}_2)$$

Let $\mathcal{F}$ and $\mathcal{H}$ be two Π_1-incompleteness witnesses. We say that $\mathcal{F}$ *is derived from* $\mathcal{H}$, or $\mathcal{H}$ *delivers* $\mathcal{F}$ (*in the standard way*) denoted $\mathcal{F} \preccurlyeq \mathcal{H}$ when for

every system $\boldsymbol{\tau}$ we have $\mathcal{T}\hbar_{\boldsymbol{\tau}} \vdash \mathcal{F}_{\boldsymbol{\tau}} \to \mathcal{H}_{\boldsymbol{\tau}}$. We note that then the unprovability of $\mathcal{F}_{\boldsymbol{\tau}}$ (in $\mathcal{T}\hbar_{\boldsymbol{\tau}}$) follows from the unprovability of $\mathcal{H}_{\boldsymbol{\tau}}$.

Some results of this paper are summarized in the following table:

Π_1-incompleteness	Does deliver $\mathfrak{G}_2$	Is derived from $\mathfrak{G}_2$
Gödel$_1$ (1931) [9]	✓	✓
Kleene$_1$ (1936) [15]	✓	✓
Rosser (1936) [22]	✓	✗
Kleene$_2$ (1950) [16]	✓	✗
Chaitin (1970) [7]	✗	✗
Boolos (1989) [5]	✗	✓

Let us denote Gödel's (respectively, Kleene's) first Π_1-incompleteness witness by $\mathfrak{G}$ (respectively, $\mathbb{IK}$); let $\overline{\mathbb{IR}}$ (respectively, $\overline{\mathbb{IK}'}$) denote an alternative version of Rosser's (respectively, Keelen's second) Π_1-incompleteness witness (which will be rigorously determined later). Let $\mathfrak{C}$ denote (one of the infinitely many Π_1-sentences that) Chaitin's incompleteness witness (proves to be true and unprovable in consistent systems). Finally, we denote by $\widetilde{\mathbb{IB}}$ a Π_1-incompleteness witness that is very similar to that of Boolos, but substantially different from his original formulation. Some results of our comparison are depicted in the following diagram (where $\mathcal{F} \cong \mathcal{H}$ abbreviates $\mathcal{F} \preccurlyeq \mathcal{H} \preccurlyeq \mathcal{F}$ and $\mathcal{F} \gtrapprox \mathcal{H}$ abbreviates $\mathcal{F} \preccurlyeq \mathcal{H} \not\preccurlyeq \mathcal{F}$):

$$\mathfrak{C} \gtrapprox \widetilde{\mathbb{IB}} \gtrapprox \mathfrak{G}_2 \cong \mathbb{IK} \cong \mathfrak{G} \gtrapprox \overline{\mathbb{IR}} \cong \overline{\mathbb{IK}'}.$$

§2. Proofs of Gödel, Rosser, and Kleene. For $n \in \mathbb{N}$, let $\overline{n}$ denote its numeral (the term representing n in the language of arithmetic). A function $f : \mathbb{N} \to \mathbb{N}$ is said to be *strongly representable* in $\mathfrak{B}$ when for a formula $\langle\langle f(x) = y \rangle\rangle$ in its language, with the only free variables x and y, we have

$$\text{for all } m,n \in \mathbb{N}, \text{ if } f(m) = n, \text{ then } \mathfrak{B} \vdash \forall y\, [\langle\langle f(\overline{m}) = y \rangle\rangle \leftrightarrow y = \overline{n}].$$

The derivability conditions are the following (for a system $\boldsymbol{\tau}$):

($\mathtt{C}_1$) $\mathfrak{B} \vdash \mathtt{Pr}_{\boldsymbol{\tau}}(\ulcorner \theta \urcorner)$ if and only if $\mathcal{T}\hbar_{\boldsymbol{\tau}} \vdash \theta$, for all sentences θ.
($\mathtt{C}_2$) $\mathfrak{B} \vdash \mathtt{Pr}_{\boldsymbol{\tau}}(\ulcorner \theta \to \eta \urcorner) \to [\mathtt{Pr}_{\boldsymbol{\tau}}(\ulcorner \theta \urcorner) \to \mathtt{Pr}_{\boldsymbol{\tau}}(\ulcorner \eta \urcorner)]$, for all sentences θ and η.
($\mathtt{C}_3$) $\mathfrak{B} \vdash \sigma \to \mathtt{Pr}_{\boldsymbol{\tau}}(\ulcorner \sigma \urcorner)$, for all Σ_1-formulas σ.

Let us start with a straightforward observation about the consistency statements:

LEMMA 2.1. *For a system $\boldsymbol{\tau}$ and a sentence θ, we have*

(1) $\mathfrak{B} \vdash \neg \mathtt{Con}_{\boldsymbol{\tau}} \to \mathtt{Pr}_{\boldsymbol{\tau}}(\ulcorner \theta \urcorner)$ *and*
(2) $\mathfrak{B} \vdash \mathtt{Pr}_{\boldsymbol{\tau}}(\ulcorner \neg \theta \urcorner) \wedge \mathtt{Pr}_{\boldsymbol{\tau}}(\ulcorner \theta \urcorner) \to \neg \mathtt{Con}_{\boldsymbol{\tau}}$.

PROOF. Both parts of the lemma follow from the derivability conditions by using the tautologies $0 \neq 0 \to \theta$ for (1), and $\neg\theta \to (\theta \to 0 \neq 0)$ for (2). ⊣

In the following proposition, we show some necessary and sufficient conditions for delivering $\mathbb{G}_2$, and being derivable from $\mathbb{G}_2$, in the standard way (cf. [19]):

PROPOSITION 2.2. *For every system τ and every Π_1-sentence ψ, we have*

(1) $\mathcal{TH}_\tau \vdash \mathtt{Con}_\tau \to \psi$ if and only if $\mathcal{TH}_\tau \vdash \neg\mathtt{Pr}_\tau(\ulcorner\psi\urcorner) \to \psi$;
(2) $\mathcal{TH}_\tau \vdash \psi \to \mathtt{Con}_\tau$ if and only if $\mathcal{TH}_\tau \vdash \psi \to \neg\mathtt{Pr}_\tau(\ulcorner\psi\urcorner)$.

PROOF. $(1_\Rightarrow)$: Suppose that $\mathcal{TH}_\tau \vdash \mathtt{Con}_\tau \to \psi$ holds. By Lemma 2.1(1) we have $\mathcal{TH}_\tau \vdash \neg\mathtt{Pr}_\tau(\ulcorner\psi\urcorner) \to \mathtt{Con}_\tau$; so $\mathcal{TH}_\tau \vdash \neg\mathtt{Pr}_\tau(\ulcorner\psi\urcorner) \to \psi$ follows.
$(1_\Leftarrow)$: Now, suppose $\mathcal{TH}_\tau \vdash \neg\mathtt{Pr}_\tau(\ulcorner\psi\urcorner) \to \psi$; so, $\mathcal{TH}_\tau \vdash \neg\psi \to \mathtt{Pr}_\tau(\ulcorner\psi\urcorner)$ holds. Also, since $\neg\psi$ is Σ_1, by (C_3), we have $\mathcal{TH}_\tau \vdash \neg\psi \to \mathtt{Pr}_\tau(\ulcorner\neg\psi\urcorner)$. Thus, by Lemma 2.1(2) we get $\mathcal{TH}_\tau \vdash \neg\psi \to \neg\mathtt{Con}_\tau$, and so $\mathcal{TH}_\tau \vdash \mathtt{Con}_\tau \to \psi$.
$(2_\Rightarrow)$: Suppose that we have $\mathcal{TH}_\tau \vdash \psi \to \mathtt{Con}_\tau$; then by (C_1) and (C_2), it follows that $\mathcal{TH}_\tau \vdash \neg\mathtt{Pr}_\tau(\ulcorner\mathtt{Con}_\tau\urcorner) \to \neg\mathtt{Pr}_\tau(\ulcorner\psi\urcorner)$. Now, by Löb's theorem (or, formalized $\mathbb{G}_2$, see [6]) we have $\mathcal{TH}_\tau \vdash \mathtt{Con}_\tau \to \neg\mathtt{Pr}_\tau(\ulcorner\mathtt{Con}_\tau\urcorner)$. Therefore, we have the desired conclusion $\mathcal{TH}_\tau \vdash \psi \to \neg\mathtt{Pr}_\tau(\ulcorner\psi\urcorner)$.
$(2_\Leftarrow)$: Now, suppose that $\mathcal{TH}_\tau \vdash \psi \to \neg\mathtt{Pr}_\tau(\ulcorner\psi\urcorner)$ holds. Then by Lemma 2.1(1), that $\mathcal{TH}_\tau \vdash \neg\mathtt{Pr}_\tau(\ulcorner\psi\urcorner) \to \mathtt{Con}_\tau$, we get $\mathcal{TH}_\tau \vdash \psi \to \mathtt{Con}_\tau$. ⊣

Let us note that the assumption "ψ is a Π_1-sentence" is used only in the proof of $(1_\Leftarrow)$: if $\mathcal{TH}_\tau \vdash \neg\mathtt{Pr}_\tau(\ulcorner\psi\urcorner) \to \psi$, then $\mathcal{TH}_\tau \vdash \mathtt{Con}_\tau \to \psi$. The rest of the implications in Proposition 2.2 hold for arbitrary ψ. That ψ should be Π_1 in $(1_\Leftarrow)$ can be seen by the following example: Let τ be a system such that $\mathcal{TH}_\tau \vdash \mathtt{Pr}_\tau(\neg\mathtt{Con}_\tau)$ and $\mathcal{TH}_\tau \nvdash \neg\mathtt{Con}_\tau$; one can take $\tau(x)$ to be the system $\varsigma(x) \vee [x = \mathtt{Pr}_\varsigma(\neg\mathtt{Con}_\varsigma)]$ for a system ς with $\mathcal{TH}_\varsigma \nvdash \neg\mathtt{Con}_\varsigma$, see [6] or [12, Theorem 36]. Then, for the Σ_1-sentence $\psi = \neg\mathtt{Con}_\tau$ we have $\mathcal{TH}_\tau \vdash \neg\mathtt{Pr}_\tau(\ulcorner\psi\urcorner) \to \psi$, but $\mathcal{TH}_\tau \nvdash \mathtt{Con}_\tau \to \psi$.

2.1. Gödel's proof. The proof of Gödel constructs a Π_1-sentence γ, for a given system τ, such that $(*)$ $\mathfrak{B} \vdash \gamma \leftrightarrow \neg\mathtt{Pr}_\tau(\ulcorner\gamma\urcorner)$ holds. If $\mathcal{TH}_\tau$ is consistent, then the unprovability of γ from $\mathcal{TH}_\tau$ follows from the derivability conditions (if $\mathcal{TH}_\tau \vdash \gamma$, then $\mathcal{TH}_\tau \vdash \mathtt{Pr}_\tau(\ulcorner\gamma\urcorner)$ by (C_1), and also $\mathcal{TH}_\tau \vdash \neg\mathtt{Pr}_\tau(\ulcorner\gamma\urcorner)$ by $(*)$, contradicting the consistency of $\mathcal{TH}_\tau$), and the truth of γ follows from the soundness of $\mathfrak{B}$ (since we have $\mathbb{N} \vDash \gamma \leftrightarrow \neg\mathtt{Pr}_\tau(\ulcorner\gamma\urcorner)$ by $(*)$, and $\mathbb{N} \vDash \neg\mathtt{Pr}_\tau(\ulcorner\gamma\urcorner)$ by $\mathcal{TH}_\tau \nvdash \gamma$, then we have $\mathbb{N} \vDash \gamma$). Proposition 2.2 immediately implies that $\mathcal{TH}_\tau \vdash \mathtt{Con}_\tau \leftrightarrow \gamma$ holds for every Π_1-sentence γ that satisfies $(*)$. Essentially, the same proof can show that $(\ddagger)$ $\mathfrak{B} \vdash \mathtt{Con}_\tau \leftrightarrow \gamma$ holds for all such sentences, and conversely, that if $(\ddagger)$ holds, then $(*)$ holds too.

THEOREM 2.3. *For every system τ and Π_1-sentence γ we have*

$$\mathfrak{B} \vdash \gamma \leftrightarrow \neg\mathtt{Pr}_\tau(\ulcorner\gamma\urcorner) \quad \textit{if and only if} \quad \mathfrak{B} \vdash \mathtt{Con}_\tau \leftrightarrow \gamma. \qquad \dashv$$

This theorem is a special case of the theorem(s) of D. de Jongh, G. Sambin, and C. Bernardi (see [6, Chapter 8]).

So, we can define *the* Gödel sentence $\mathbb{G}_\tau$ of a system τ to be any of the Π_1-sentences γ that satisfy $\mathfrak{B} \vdash \gamma \leftrightarrow \neg\mathtt{Pr}_\tau(\ulcorner\gamma\urcorner)$; noting that $\mathbb{G}_\tau$ is unique up to equivalence, even provably so in $\mathfrak{B}$.

2.2. Rosser's proof. The proof of Rosser constructs, by using the diagonal lemma, a Π_1-sentence ρ, for a given system $\boldsymbol{\tau}$, such that

$$\mathfrak{B} \vdash \rho \leftrightarrow \forall x[\mathtt{prf}_{\boldsymbol{\tau}}(x,\ulcorner\rho\urcorner) \rightarrow \exists y < x\, \mathtt{prf}_{\boldsymbol{\tau}}(y,\ulcorner\neg\rho\urcorner)]$$

holds. Let us call any such Π_1-sentence ρ, a *Rosserian* sentence of the system $\boldsymbol{\tau}$.

The classical proof of Rosser shows the independence of ρ from $\mathcal{TH}_{\boldsymbol{\tau}}$, if $\mathcal{TH}_{\boldsymbol{\tau}}$ is (only) consistent (and not necessarily Σ_1-sound): if $\mathcal{TH}_{\boldsymbol{\tau}} \vdash \rho$, then $\mathcal{TH}_{\boldsymbol{\tau}} \vdash \forall x[\mathtt{prf}_{\boldsymbol{\tau}}(x,\ulcorner\rho\urcorner) \rightarrow \exists y < x\, \mathtt{prf}_{\boldsymbol{\tau}}(y,\ulcorner\neg\rho\urcorner)]$, and also, by $(\mathtt{C}_1)$, $\mathcal{TH}_{\boldsymbol{\tau}} \vdash \mathtt{prf}_{\boldsymbol{\tau}}(\overline{n},\ulcorner\rho\urcorner)$ for some $n \in \mathbb{N}$; whence, $\mathcal{TH}_{\boldsymbol{\tau}} \vdash \exists y < \overline{n}\, \mathtt{prf}_{\boldsymbol{\tau}}(y,\ulcorner\neg\rho\urcorner)$, but the Δ_0-sentence $\exists y < \overline{n}\, \mathtt{prf}_{\boldsymbol{\tau}}(y,\ulcorner\neg\rho\urcorner)$ would be false by the consistency of $\mathcal{TH}_{\boldsymbol{\tau}}$, and so would be refutable in $\mathfrak{B}$, contradiction. Also, the assumption $\mathcal{TH}_{\boldsymbol{\tau}} \vdash \neg\rho$, would imply $\mathcal{TH}_{\boldsymbol{\tau}} \vdash \exists x[\mathtt{prf}_{\boldsymbol{\tau}}(x,\ulcorner\rho\urcorner) \wedge \forall y < x \neg\mathtt{prf}_{\boldsymbol{\tau}}(y,\ulcorner\neg\rho\urcorner)]$ on the one hand, and $\mathfrak{B} \vdash \mathtt{prf}_{\boldsymbol{\tau}}(\overline{m},\ulcorner\neg\rho\urcorner)$ for some $m \in \mathbb{N}$, by (C_1), on the other hand; thus, $\mathcal{TH}_{\boldsymbol{\tau}} \vdash \exists x \leqslant \overline{m}\, \mathtt{prf}_{\boldsymbol{\tau}}(x,\ulcorner\rho\urcorner)$, but by the consistency of $\mathcal{TH}_{\boldsymbol{\tau}}$, the Δ_0-sentence $\exists x \leqslant \overline{m}\, \mathtt{prf}_{\boldsymbol{\tau}}(x,\ulcorner\rho\urcorner)$ would be false, and so refutable in $\mathfrak{B}$, contradiction.

Rosser's theorem is not derivable from $\mathbb{G}_2$ in the standard way (though, it does deliver $\mathbb{G}_2$): For a consistent system $\boldsymbol{\tau}$, let $\boldsymbol{\varrho}(x) = \boldsymbol{\tau}(x) \vee [x = \neg\mathtt{Con}_{\boldsymbol{\tau}}]$; then $\mathcal{TH}_{\boldsymbol{\varrho}}$ is consistent by $\mathbb{G}_2$, and $\mathcal{TH}_{\boldsymbol{\varrho}} \vdash \neg\mathtt{Con}_{\boldsymbol{\varrho}}$. So, for every Rosserian sentence ρ of $\boldsymbol{\varrho}$, we have $\mathcal{TH}_{\boldsymbol{\varrho}} \nvdash \rho \rightarrow \mathtt{Con}_{\boldsymbol{\varrho}}$, since otherwise $\mathcal{TH}_{\boldsymbol{\varrho}} \vdash \neg\rho$ would hold, contradicting Rosser's theorem. Below we show a stronger result.[1]

THEOREM 2.4. *Let $\boldsymbol{\tau}$ be a system, and ρ be a Rosserian sentence of $\boldsymbol{\tau}$. Then $\mathfrak{B} \vdash \mathtt{Con}_{\boldsymbol{\tau}} \rightarrow \rho$ holds; and if $\mathcal{TH}_{\boldsymbol{\tau}}$ is consistent, then $\mathcal{TH}_{\boldsymbol{\tau}} \nvdash \rho \rightarrow \mathtt{Con}_{\boldsymbol{\tau}}$.*

PROOF. By $\mathfrak{B} \vdash \neg\rho \leftrightarrow \exists x[\mathtt{prf}_{\boldsymbol{\tau}}(x,\ulcorner\rho\urcorner) \wedge \forall y < x\, \neg\mathtt{prf}_{\boldsymbol{\tau}}(y,\ulcorner\neg\rho\urcorner)]$, we have that $\mathfrak{B} \vdash \neg\rho \rightarrow \mathtt{Pr}_{\boldsymbol{\tau}}(\ulcorner\rho\urcorner)$. Thus, Proposition 2.2(1) implies $\mathfrak{B} \vdash \mathtt{Con}_{\boldsymbol{\tau}} \rightarrow \rho$.

For showing the second part, we first show a formalized version of Rosser's theorem, $\mathfrak{B} \vdash \mathtt{Con}_{\boldsymbol{\tau}} \rightarrow \neg\mathtt{Pr}_{\boldsymbol{\tau}}(\ulcorner\neg\rho\urcorner)$. Reason inside $\mathfrak{B} + \mathtt{Con}_{\boldsymbol{\tau}} + \mathtt{Pr}_{\boldsymbol{\tau}}(\ulcorner\neg\rho\urcorner)$:

> We have $\neg\rho \leftrightarrow \exists x[\mathtt{prf}_{\boldsymbol{\tau}}(x,\ulcorner\rho\urcorner) \wedge \forall y < x\, \neg\mathtt{prf}_{\boldsymbol{\tau}}(y,\ulcorner\neg\rho\urcorner)]$, and so $\mathtt{Pr}_{\boldsymbol{\tau}}(\ulcorner\neg\rho\urcorner)$ implies $(\ddagger)$ $\mathtt{Pr}_{\boldsymbol{\tau}}\big(\ulcorner\exists b[\mathtt{prf}_{\boldsymbol{\tau}}(b,\ulcorner\rho\urcorner) \wedge \forall y < b \neg\mathtt{prf}_{\boldsymbol{\tau}}(y,\ulcorner\neg\rho\urcorner)]\urcorner\big)$ by $(\mathtt{C}_1,\mathtt{C}_2)$. Also, $\mathtt{Pr}_{\boldsymbol{\tau}}(\ulcorner\neg\rho\urcorner)$ implies the existence of some a such that $(\ddagger\ddagger)$ $\mathtt{prf}_{\boldsymbol{\tau}}(a,\ulcorner\neg\rho\urcorner)$ holds. Now, by $(\ddagger)$, $(\ddagger\ddagger)$, and $\mathtt{Con}_{\boldsymbol{\tau}}$, we have $(*)$ $\mathtt{Pr}_{\boldsymbol{\tau}}\big(\ulcorner\exists b \leqslant a\, \mathtt{prf}_{\boldsymbol{\tau}}(b,\ulcorner\rho\urcorner)\urcorner\big)$. On the other hand, by the assumption $\mathtt{Con}_{\boldsymbol{\tau}} + \mathtt{Pr}_{\boldsymbol{\tau}}(\ulcorner\neg\rho\urcorner)$, we have $\neg\mathtt{Pr}_{\boldsymbol{\tau}}(\ulcorner\rho\urcorner)$, and so the Δ_0-formula $\forall x \leqslant a \neg\mathtt{prf}_{\boldsymbol{\tau}}(x,\ulcorner\rho\urcorner)$ is true; whence, by $(\mathtt{C}_3)$, we have $(**)$ $\mathtt{Pr}_{\boldsymbol{\tau}}\big(\ulcorner\forall x \leqslant a \neg\mathtt{prf}_{\boldsymbol{\tau}}(x,\ulcorner\rho\urcorner)\urcorner\big)$. Now, $(*)$ and $(**)$ contradict $\mathtt{Con}_{\boldsymbol{\tau}}$.

Assume now that $\mathcal{TH}_{\boldsymbol{\tau}}$ is consistent; and assume (for the sake of a contradiction) that $\mathcal{TH}_{\boldsymbol{\tau}} \vdash \rho \rightarrow \mathtt{Con}_{\boldsymbol{\tau}}$ holds. Then by $\mathcal{TH}_{\boldsymbol{\tau}} \vdash \mathtt{Con}_{\boldsymbol{\tau}} \rightarrow \neg\mathtt{Pr}_{\boldsymbol{\tau}}(\ulcorner\neg\rho\urcorner)$, proved above, we have $\mathcal{TH}_{\boldsymbol{\tau}} + \rho \vdash \neg\mathtt{Pr}_{\boldsymbol{\tau}}(\ulcorner\neg\rho\urcorner)$. This contradicts

[1] Theorem 2.4 was first proved in [18, p. 16] where it is stated that this result was "implicit in the papers of Gödel [9] and Rosser [22]" (see also the end of [17]). It is also proved in [2] (see the Lemma on page 405) in which it is stated that although this result "was certainly known before," the author "was unable to find a proof of it in the literature."

$\mathbb{G}_2$, unless $\mathcal{TH}_\tau + \rho$ is inconsistent, or $\mathcal{TH}_\tau \vdash \neg\rho$; and this contradicts Rosser's theorem, unless $\mathcal{TH}_\tau$ is inconsistent. So, we showed that if $\mathcal{TH}_\tau$ is consistent, then $\mathcal{TH}_\tau \nvdash \rho \to \mathtt{Con}_\tau$. ⊣

It follows from Theorem 2.3 that all the Π_1-sentences γ for which we have that $\mathfrak{B} \vdash \gamma \leftrightarrow \neg \mathtt{Pr}_\tau(\ulcorner\gamma\urcorner)$ are $\mathfrak{B}$-provably equivalent with (each other and with) $\mathtt{Con}_\tau$. That all the Rosserian sentences of τ, that is, all the Π_1-sentences ρ for which we have $\mathfrak{B} \vdash \rho \leftrightarrow \forall x[\mathtt{prf}_\tau(x,\ulcorner\rho\urcorner) \to \exists y < x\, \mathtt{prf}_\tau(y,\ulcorner\neg\rho\urcorner)]$, are equivalent with each other (in $\mathcal{TH}_\tau$) was posed as an open question in [18, p. 16]. This was answered in [10] as follows: there are standard proof predicates[2] for which all the Rosserian sentences are equivalent (see [10, Theorem 6.2], and [30] for a correct proof); and there are standard proof predicates for which there are nonequivalent Rosserian sentences (see [10, Theorem 6.1]). For a given system τ, let $\underline{\overline{\mathtt{prf}}}_\tau(y,x)$ be a proof predicate of τ all of whose Rosserian sentences are equivalent (see [4]). Then, (a variant of) *the* Rosser sentence $\overline{\mathbb{R}}_\tau$ of τ can be defined to be any of the Π_1-sentences ρ that satisfy $\mathfrak{B} \vdash \rho \leftrightarrow \forall x[\underline{\overline{\mathtt{prf}}}_\tau(x,\ulcorner\rho\urcorner) \to \exists y < x\, \underline{\overline{\mathtt{prf}}}_\tau(y,\ulcorner\neg\rho\urcorner)]$.

So far, we have noticed that for every system τ,

$$\mathfrak{B} \vdash \mathbb{G}_\tau \leftrightarrow \mathtt{Con}_\tau \text{ and } \mathfrak{B} \vdash \mathtt{Con}_\tau \to \overline{\mathbb{R}}_\tau; \text{ but } \mathcal{TH}_\tau \nvdash \overline{\mathbb{R}}_\tau \to \mathtt{Con}_\tau \text{ for consistent } \mathcal{TH}_\tau.$$

Let us note that $\overline{\mathbb{R}}_\tau$ is sensitive to implementation details modulo provable equivalence in the ambient theory, while $\mathbb{G}_\tau$ is not so.

2.3. Kleene's proof(s). The proofs of Gödel and Rosser use the diagonal (aka self-referential) lemma for constructing the unprovable sentences. Kleene's (both first and second) proof can be considered diagonal-free in a sense, since it does not use this lemma directly, and avoids self-referentiality. However, as will be seen below, one can still argue that there could be some (at least, hidden) circularity in the proof. For stating Kleene's proofs, let us fix the notation.

DEFINITION 2.5. Let all the *unary* partial recursive functions be effectively (recursively) listed as $\varphi_0, \varphi_1, \varphi_2, \ldots$. For $m,n \in \mathbb{N}$, if φ_m is defined at n (i.e., $\varphi_m(n)$ exists), then we write $\varphi_m(n)\!\downarrow$, and say that φ_m halts on n; likewise, $\varphi_m(n)\!\uparrow$ means that the function φ_m is not defined at n. Let us take $\langle\langle \varphi_{\overline{m}}(\overline{n})\!\uparrow\rangle\rangle$ to be the formula, in the language of arithmetic, that expresses $\varphi_m(n)\!\uparrow$. ⊣

Let us note that $\langle\langle \varphi_{\overline{m}}(\overline{n})\!\uparrow\rangle\rangle$ is a Π_1-sentence. Kleene's proof [15] shows the existence of some $k \in \mathbb{N}$, for a given system τ, such that the sentence $\langle\langle \varphi_{\overline{k}}(\overline{k})\!\uparrow\rangle\rangle$ is (true but) unprovable in $\mathcal{TH}_\tau$, if $\mathcal{TH}_\tau$ is consistent. The nonconstructive version of the proof goes as follows: since the set $\{n \in \mathbb{N} \mid \mathcal{TH}_\tau \vdash \langle\langle \varphi_{\overline{n}}(\overline{n})\!\uparrow\rangle\rangle\}$ is RE, but the set $\{n \in \mathbb{N} \mid \mathbb{N} \vDash \langle\langle \varphi_{\overline{n}}(\overline{n})\!\uparrow\rangle\rangle\}$ is not, and the former is contained in the latter for consistent system $\mathcal{TH}_\tau$, so there must exist some k such that we have $\mathbb{N} \vDash \langle\langle \varphi_{\overline{k}}(\overline{k})\!\uparrow\rangle\rangle$ but $\mathcal{TH}_\tau \nvdash \langle\langle \varphi_{\overline{k}}(\overline{k})\!\uparrow\rangle\rangle$. Indeed, any RE and (effectively) undecidable set could be used for the proof; so, the existence of an RE and undecidable set implies Gödel's first incompleteness theorem,

[2] A proof predicate $\lambda(y,x)$ is called *standard*, when its provability predicate, defined as $\Lambda(x) = \exists y\, \lambda(y,x)$, satisfies the derivability conditions.

by this argument of Kleene. It suffices for φ_k to have the property $\varphi_k(n)\downarrow \iff \mathcal{TH}_\tau \vdash \langle\langle \varphi_{\overline{n}}(\overline{n})\uparrow\rangle\rangle$, for every $n\in\mathbb{N}$ (see [25, Theorem 2.2]); it is worth noting that there are indeed infinitely many such k's. Now, the (true and) unprovable sentence of Kleene's first proof, for a given system τ, can be constructed in a diagonal-free way as follows.

DEFINITION 2.6. For a system τ, let $\mathfrak{t}\in\mathbb{N}$ be an index (out of the infinitely many indexes) for the recursive function $n\mapsto \mu z{:}\,\mathtt{prf}_\tau(z,\ulcorner\langle\langle\varphi_{\overline{n}}(\overline{n})\uparrow\rangle\rangle\urcorner)$. Let $\mathbb{K}_\tau = \langle\langle\varphi_{\overline{\mathfrak{t}}}(\overline{\mathfrak{t}})\uparrow\rangle\rangle$ be *Kleene's* (first) *sentence* for system τ. ⊣

Let us note that by the definition of $\mathfrak{t}$, $\varphi_{\mathfrak{t}}(\mathfrak{t}) = \mu z{:}\,\mathtt{prf}_\tau(z,\ulcorner\langle\langle\varphi_{\overline{\mathfrak{t}}}(\overline{\mathfrak{t}})\uparrow\rangle\rangle\urcorner)$, and so $\varphi_{\mathfrak{t}}(\mathfrak{t})\uparrow \iff \mathcal{TH}_\tau \nvdash \langle\langle\varphi_{\overline{\mathfrak{t}}}(\overline{\mathfrak{t}})\uparrow\rangle\rangle$, or equivalently $\mathbb{K}_\tau \leftrightarrow \neg\mathtt{Pr}_\tau(\ulcorner\mathbb{K}_\tau\urcorner)$ (even provably in $\mathfrak{B}$), which resembles Gödel's equivalence $\mathbb{G}_\tau \leftrightarrow \neg\mathtt{Pr}_\tau(\ulcorner\mathbb{G}_\tau\urcorner)$. So, the equivalence of Kleene's first incompleteness witness with $\mathbb{G}_2$ is a consequence of Theorem 2.3 (which also shows that $\mathbb{K}_\tau$ is not sensitive to various implementation details):

COROLLARY 2.7. *For every system* τ *we have* $\mathfrak{B}\vdash \mathtt{Con}_\tau \leftrightarrow \mathbb{K}_\tau$. ⊣

2.4. Kleene's symmetric proof. Kleene's first sentence is not independent from the system if the system is consistent (see [25, Theorem 2.3]); and so Kleene [16] gave another proof for the Gödel–Rosser theorem, which was called by him "a symmetric form" of Gödel's (incompleteness) theorem; see also [23, Theorem 14].

DEFINITION 2.8. Let us effectively list all the *binary* partial recursive functions as $\phi_0,\phi_1,\phi_2,\ldots$. Let $\phi_n(k,l)\downarrow_m$ mean that the binary partial recursive function ϕ_n is defined at (k,l) and its value can be computed (in a fixed programming language) in m steps (or less). As before, its formalization in the language of arithmetic is denoted by $\langle\langle\phi_{\overline{n}}(\overline{k},\overline{l})\downarrow_{\overline{m}}\rangle\rangle$. ⊣

Let us note that $\langle\langle\phi_x(u,v)\downarrow_y\rangle\rangle$ can be written by a Δ_0-formula of the free variables u,v,x, and y. For a system τ, let $\mathfrak{f}_\tau:\mathbb{N}\to\mathbb{N}$ be the τ-proof search function, $\mathfrak{f}_\tau(u) = \mu z{:}\,\overline{\mathtt{prf}}_\tau(z,u)$, whose algorithm is as follows:

> input u, put $i:=0$, and run the sub-program $\sharp_i$;
> $\sharp_i$: check if i is a τ-proof (if for some $j\leqslant i$, $\overline{\mathtt{prf}}_\tau(i,j)$ holds); if not, then put $i:=i+1$ and repeat $\sharp_i$; if yes, then compute what i proves (the above j); if $j\neq u$, then put $i:=i+1$ and repeat $\sharp_i$; if $j=u$, then output i and halt.

DEFINITION 2.9. For numbers $m,n\in\mathbb{N}$, let $\mathfrak{I}^m_n$ be the following sentence: $\forall x[\langle\langle\phi_{\overline{m}}(\overline{m},\overline{n})\downarrow_x\rangle\rangle \to \exists y<x\langle\langle\phi_{\overline{n}}(\overline{m},\overline{n})\downarrow_y\rangle\rangle]$. For a given system τ, let r and s be some indexes for the (binary recursive) functions $(m,n)\mapsto \mathfrak{f}_\tau(\ulcorner\mathfrak{I}^m_n\urcorner)$ and $(m,n)\mapsto \mathfrak{f}_\tau(\ulcorner\neg\mathfrak{I}^m_n\urcorner)$, respectively. For a given system τ, let $\overline{\mathbb{K}'}_\tau = \mathfrak{I}^{\mathfrak{r}}_{\mathfrak{s}}$ be *Kleene's second sentence* for τ. ⊣

Thus, for a system τ, $\phi_{\mathfrak{r}}(m,n) = \mathfrak{f}_\tau(\ulcorner\mathfrak{I}^m_n\urcorner)$ and $\phi_{\mathfrak{s}}(m,n) = \mathfrak{f}_\tau(\ulcorner\neg\mathfrak{I}^m_n\urcorner)$; also $\overline{\mathbb{K}'}_\tau = \forall x[\langle\langle\phi_{\overline{\mathfrak{r}}}(\overline{\mathfrak{r}},\overline{\mathfrak{s}})\downarrow_x\rangle\rangle \to \exists y<x\langle\langle\phi_{\overline{\mathfrak{s}}}(\overline{\mathfrak{r}},\overline{\mathfrak{s}})\downarrow_y\rangle\rangle]$, which is a Π_1-sentence.

For a consistent system τ, the independence of $\overline{\mathbb{K}'}_\tau$ from $\mathcal{TH}_\tau$ can be shown along the lines of Rosser's proof: If $\mathcal{TH}_\tau\vdash\overline{\mathbb{K}'}_\tau$, then $\phi_{\mathfrak{r}}(\mathfrak{r},\mathfrak{s})\downarrow_m$ for some $m\in\mathbb{N}$,

and so $\mathcal{T}\hbar_{\boldsymbol{\tau}} \vdash \exists y < \overline{m}\langle\langle \phi_{\bar{\mathfrak{s}}}(\bar{\mathfrak{r}},\bar{\mathfrak{s}})\downarrow_y\rangle\rangle$. But by the consistency of $\mathcal{T}\hbar_{\boldsymbol{\tau}}$ we have $\mathcal{T}\hbar_{\boldsymbol{\tau}} \nvdash \neg\overline{\mathbb{IK}'}_{\boldsymbol{\tau}}$ and so $\phi_{\bar{\mathfrak{s}}}(\bar{\mathfrak{r}},\bar{\mathfrak{s}})\uparrow$; whence, the Δ_0-sentence $\exists y < \overline{m}\langle\langle \phi_{\bar{\mathfrak{s}}}(\bar{\mathfrak{r}},\bar{\mathfrak{s}})\downarrow_y\rangle\rangle$ is false and so should be refutable in $\mathcal{T}\hbar_{\boldsymbol{\tau}}$, contradiction. Also, if $\mathcal{T}\hbar_{\boldsymbol{\tau}} \vdash \neg\overline{\mathbb{IK}'}_{\boldsymbol{\tau}}$, then $\phi_{\mathfrak{s}}(\mathfrak{r},\mathfrak{s})\downarrow_n$ for some $n \in \mathbb{N}$, and so $\mathcal{T}\hbar_{\boldsymbol{\tau}} \vdash \exists x \leqslant \overline{n}\langle\langle \phi_{\bar{\mathfrak{r}}}(\bar{\mathfrak{r}},\bar{\mathfrak{s}})\downarrow_x\rangle\rangle$. But the Δ_0-sentence $\exists x \leqslant \overline{n}\langle\langle \phi_{\bar{\mathfrak{r}}}(\bar{\mathfrak{r}},\bar{\mathfrak{s}})\downarrow_x\rangle\rangle$ is false by the consistency of $\mathcal{T}\hbar_{\boldsymbol{\tau}}$ (which implies $\mathcal{T}\hbar_{\boldsymbol{\tau}} \nvdash \overline{\mathbb{IK}'}_{\boldsymbol{\tau}}$, thus $\phi_{\bar{\mathfrak{r}}}(\bar{\mathfrak{r}},\bar{\mathfrak{s}})\uparrow$) and so it should be refutable in $\mathcal{T}\hbar_{\boldsymbol{\tau}}$; contradiction again.

Also, very similarly to the proof of Theorem 2.4, it can be shown that for every system $\boldsymbol{\tau}$ we have $\mathcal{T}\hbar_{\boldsymbol{\tau}} \vdash \mathtt{Con}_{\boldsymbol{\tau}} \to \overline{\mathbb{IK}'}_{\boldsymbol{\tau}}$; and $\mathcal{T}\hbar_{\boldsymbol{\tau}} \nvdash \overline{\mathbb{IK}'}_{\boldsymbol{\tau}} \to \mathtt{Con}_{\boldsymbol{\tau}}$ if $\mathcal{T}\hbar_{\boldsymbol{\tau}}$ is consistent. We show, more generally, that $\overline{\mathbb{IK}'}_{\boldsymbol{\tau}}$ is a Rosserian sentence of $\boldsymbol{\tau}$, and so this follows directly from Theorem 2.4.

If $x = \mathfrak{f}_{\boldsymbol{\tau}}(u)$ for a system $\boldsymbol{\tau}$, and some u, then let $\widehat{x}$ be the least s such that $\mathfrak{f}_{\boldsymbol{\tau}}(u)\downarrow_s$ holds. We note that, for every u,v,x,y, if $\mathfrak{f}_{\boldsymbol{\tau}}(u) = x < y = \mathfrak{f}_{\boldsymbol{\tau}}(v)$, then $\widehat{x} < \widehat{y}$: this is because the algorithm of $\mathfrak{f}_{\boldsymbol{\tau}}$, on the input v, has already checked x, before halting at the step $\widehat{y}$, to see if it is a $\boldsymbol{\tau}$-proof, and if (yes, then) it is a $\boldsymbol{\tau}$-proof of v. The algorithm has noticed that x is a $\boldsymbol{\tau}$-proof of some $u \neq v$, and then has gone to the next step (to check $x+1$ and so on). So, the number of steps that $\mathfrak{f}_{\boldsymbol{\tau}}$ needs to calculate y (on the input v) is greater than the number of steps that it needs to calculate x (on the input u); thus $\widehat{x} < \widehat{y}$.

THEOREM 2.10. *For every system* $\boldsymbol{\tau}$, *the sentence* $\overline{\mathbb{IK}'}_{\boldsymbol{\tau}}$ *of* $\boldsymbol{\tau}$ *is a Rosserian sentence*: $\mathfrak{B} \vdash \overline{\mathbb{IK}'}_{\boldsymbol{\tau}} \leftrightarrow \forall x[\underline{\overline{\mathtt{prf}}}_{\boldsymbol{\tau}}(x, \ulcorner\overline{\mathbb{IK}'}_{\boldsymbol{\tau}}\urcorner) \to \exists y < x\, \underline{\overline{\mathtt{prf}}}_{\boldsymbol{\tau}}(y, \ulcorner\overline{\mathbb{IK}'}_{\boldsymbol{\tau}}\urcorner)]$.

PROOF. Reason inside $\mathfrak{B}$:

Suppose that $\overline{\mathbb{IK}'}_{\boldsymbol{\tau}}$ holds, and for some a, we have $\underline{\overline{\mathtt{prf}}}_{\boldsymbol{\tau}}(a, \ulcorner\overline{\mathbb{IK}'}_{\boldsymbol{\tau}}\urcorner)$. Let α be the minimum x with $\underline{\overline{\mathtt{prf}}}_{\boldsymbol{\tau}}(x, \ulcorner\overline{\mathbb{IK}'}_{\boldsymbol{\tau}}\urcorner)$; then $a \geqslant \alpha$, and $\langle\langle \phi_{\bar{\mathfrak{r}}}(\bar{\mathfrak{r}},\bar{\mathfrak{s}})\downarrow_{\widehat{\alpha}}\rangle\rangle$ holds. Thus, by $\overline{\mathbb{IK}'}_{\boldsymbol{\tau}}$, there exists some $\beta < \widehat{\alpha}$ such that $\langle\langle \phi_{\bar{\mathfrak{s}}}(\bar{\mathfrak{r}},\bar{\mathfrak{s}})\downarrow_\beta\rangle\rangle$ holds. So, for $b = \phi_{\bar{\mathfrak{s}}}(\bar{\mathfrak{r}},\bar{\mathfrak{s}})$ we have $\underline{\overline{\mathtt{prf}}}_{\boldsymbol{\tau}}(b, \ulcorner\neg\overline{\mathbb{IK}'}_{\boldsymbol{\tau}}\urcorner)$. We show that $b < a$; if on the contrary $b \geqslant a$, then we would have $b \geqslant \alpha$, and so $\beta \geqslant \widehat{b} \geqslant \widehat{\alpha}$; contradiction with $\beta < \widehat{\alpha}$. Thus, if $\overline{\mathbb{IK}'}_{\boldsymbol{\tau}}$ holds, then for all a with $\underline{\overline{\mathtt{prf}}}_{\boldsymbol{\tau}}(a, \ulcorner\overline{\mathbb{IK}'}_{\boldsymbol{\tau}}\urcorner)$ there exists some $b < a$ with $\underline{\overline{\mathtt{prf}}}_{\boldsymbol{\tau}}(b, \ulcorner\neg\overline{\mathbb{IK}'}_{\boldsymbol{\tau}}\urcorner)$.

Now, suppose that $\neg\overline{\mathbb{IK}'}_{\boldsymbol{\tau}}$ holds. So, there exists some p such that $\langle\langle \phi_{\bar{\mathfrak{r}}}(\bar{\mathfrak{r}},\bar{\mathfrak{s}})\downarrow_p \rangle\rangle$ holds, and for no $q < p$ can $\langle\langle \phi_{\bar{\mathfrak{s}}}(\bar{\mathfrak{r}},\bar{\mathfrak{s}})\downarrow_q\rangle\rangle$ hold. Then, let $a = \phi_{\bar{\mathfrak{r}}}(\bar{\mathfrak{r}},\bar{\mathfrak{s}})$; so, $p \geqslant \widehat{a}$, and we have $\underline{\overline{\mathtt{prf}}}_{\boldsymbol{\tau}}(a, \ulcorner\overline{\mathbb{IK}'}_{\boldsymbol{\tau}}\urcorner)$. We show that for no $b < a$ can $\underline{\overline{\mathtt{prf}}}_{\boldsymbol{\tau}}(b, \ulcorner\neg\overline{\mathbb{IK}'}_{\boldsymbol{\tau}}\urcorner)$ hold. Assume, on the contrary, that for some $b < a$ we have $\underline{\overline{\mathtt{prf}}}_{\boldsymbol{\tau}}(b, \ulcorner\neg\overline{\mathbb{IK}'}_{\boldsymbol{\tau}}\urcorner)$. Then, $\langle\langle \phi_{\bar{\mathfrak{s}}}(\bar{\mathfrak{r}},\bar{\mathfrak{s}})\downarrow\rangle\rangle$; let $d = \phi_{\bar{\mathfrak{s}}}(\bar{\mathfrak{r}},\bar{\mathfrak{s}})$. So, we have $\langle\langle \phi_{\bar{\mathfrak{s}}}(\bar{\mathfrak{r}},\bar{\mathfrak{s}})\downarrow_{\widehat{d}}\rangle\rangle$, and we also have $d \leqslant b < a$. Thus, $d < a$ holds, and so we have $\widehat{d} < \widehat{a} \leqslant p$; this is a contradiction with the property of p (that $\forall q < p\ \neg\langle\langle \phi_{\bar{\mathfrak{s}}}(\bar{\mathfrak{r}},\bar{\mathfrak{s}})\downarrow_q\rangle\rangle$). Thus, if $\neg\overline{\mathbb{IK}'}_{\boldsymbol{\tau}}$ holds, then there exists some a such that $\underline{\overline{\mathtt{prf}}}_{\boldsymbol{\tau}}(a, \ulcorner\overline{\mathbb{IK}'}_{\boldsymbol{\tau}}\urcorner) \wedge \forall b < a\ \neg\underline{\overline{\mathtt{prf}}}_{\boldsymbol{\tau}}(b, \ulcorner\neg\overline{\mathbb{IK}'}_{\boldsymbol{\tau}}\urcorner)$. ⊣

Therefore, under some certain considerations (and the choice of $\underline{\overline{\mathtt{prf}}}_{\boldsymbol{\tau}}(y,x)$ for proof predicate), Kleene's second Π_1-incompleteness witness is equivalent to Rosser's Π_1-incompleteness witness.

COROLLARY 2.11. *For every system* $\boldsymbol{\tau}$, *we have* $\mathfrak{B} \vdash \overline{\mathbb{K}'}_{\boldsymbol{\tau}} \leftrightarrow \overline{\mathbb{R}}_{\boldsymbol{\tau}}$. *So, we have* $\mathfrak{B} \vdash \mathtt{Con}_{\boldsymbol{\tau}} \rightarrow \overline{\mathbb{K}'}_{\boldsymbol{\tau}}$; *and* $\mathcal{T}\!\mathit{h}_{\boldsymbol{\tau}} \nvdash \overline{\mathbb{K}'}_{\boldsymbol{\tau}} \rightarrow \mathtt{Con}_{\boldsymbol{\tau}}$ *if* $\mathcal{T}\!\mathit{h}_{\boldsymbol{\tau}}$ *is consistent* ⊣

Thus far, we have shown that Gödel's first theorem is equivalent with $\mathbb{G}_2$ and with Kleene's first theorem; (a variant of) Rosser's theorem is equivalent with (a variant of) Kleene's second theorem, and it does deliver $\mathbb{G}_2$, but is dot derivable from $\mathbb{G}_2$. In picture:

$$(\mathbb{G}\mathtt{ödel}_2 \approx \mathbb{K}\mathtt{leene} \approx \mathbb{G}\mathtt{ödel}_1) \mathrel{\overset{\sim}{\approx}} (\overline{\mathbb{R}}\mathtt{osser} \approx \overline{\mathbb{K}'}\mathtt{leene}).$$

§3. Proofs of Chaitin and Boolos. We say that the Π_1-incompleteness witness $\mathcal{F}$ is *constructive* (or *effective*), when $\mathcal{F}_{\boldsymbol{\tau}}$ can be effectively (computably) constructed form $\boldsymbol{\tau}$. We say that the Π_1-incompleteness witness $\mathcal{F}$ has *the Rosser property*, when $\mathcal{F}_{\boldsymbol{\tau}}$ is independent from $\mathcal{T}\!\mathit{h}_{\boldsymbol{\tau}}$, for every consistent system $\boldsymbol{\tau}$. Constructivity and the Rosser property of the Π_1-incompleteness theorems of Gödel, Rosser, Kleene, Chaitin, and Boolos were studied in [25] (see the table on its p. 579). There, it was shown that the Π_1-incompleteness theorems of Gödel, Rosser, and (both theorems of) Kleene are constructive, and the theorems of Chaitin and Boolos are not; also none of the theorems of Gödel, Kleene's first, or Boolos have the Rosser property, while the theorems of Rosser, Kleene's second, and (a variant of) Chaitin do have the Rosser property.

3.1. Chaitin's proof. There are several variants of Chaitin's theorem. Here, we consider one of the simple ones.

DEFINITION 3.1. The *Kolmogorov–Chaitin Complexity* function is defined to be the mapping $\mathcal{K}(w) = \boldsymbol{\mu} e : [\boldsymbol{\varphi}_e(0) = w]$, for $w \in \mathbb{N}$. Let $\langle\langle \mathcal{K}(x) > y \rangle\rangle$ denote the Π_1-formula $\forall v \leqslant y[\langle\langle \boldsymbol{\varphi}_v(0)\downarrow \rangle\rangle \rightarrow \langle\langle \boldsymbol{\varphi}_v(0) \neq x \rangle\rangle]$, in the language of arithmetic, with the free variables x and y, that expresses $\mathcal{K}(x) > y$. ⊣

Chaitin's theorem [7] shows the existence of some $\mathfrak{c}_{\boldsymbol{\tau}} \in \mathbb{N}$, for a given system $\boldsymbol{\tau}$, such that for every $w, e \in \mathbb{N}$ with $e \geqslant \mathfrak{c}_{\boldsymbol{\tau}}$ we have $\mathcal{T}\!\mathit{h}_{\boldsymbol{\tau}} \nvdash \langle\langle \mathcal{K}(\overline{w}) > \overline{e} \rangle\rangle$ if $\mathcal{T}\!\mathit{h}_{\boldsymbol{\tau}}$ is consistent. It suffices to put for a system $\boldsymbol{\tau}$,

$$\boldsymbol{\varphi}_{\mathfrak{c}_{\boldsymbol{\tau}}}(x) = \boldsymbol{\pi}_1\big[\boldsymbol{\mu} z : \mathtt{prf}_{\boldsymbol{\tau}}\big(\boldsymbol{\pi}_2[z], \ulcorner \langle\langle \mathcal{K}(\boldsymbol{\pi}_1[z]) > x + \mathfrak{c}_{\boldsymbol{\tau}} \rangle\rangle \urcorner\big)\big],$$

which is possible by Kleene's recursion (or fixed point) theorem; here, $\boldsymbol{\pi}_1, \boldsymbol{\pi}_2$ are the projection functions of a fixed enumeration of ordered pairs (i.e., if we have a bijective mapping $(a,b) \mapsto \langle a,b \rangle$ between ordered pairs of numbers and numbers, then we have $\boldsymbol{\pi}_1(\langle a,b \rangle) = a$ and $\boldsymbol{\pi}_2(\langle a,b \rangle) = b$).

Chaitin's proof goes as follows: If $\mathcal{T}\!\mathit{h}_{\boldsymbol{\tau}}$ is consistent, and $\mathcal{T}\!\mathit{h}_{\boldsymbol{\tau}} \vdash \langle\langle \mathcal{K}(\overline{w}) > \overline{e} \rangle\rangle$ for some $w, e \in \mathbb{N}$ with $e \geqslant \mathfrak{c}_{\boldsymbol{\tau}}$, then let $z = \langle u, p \rangle$ be the minimum ordered pair such that $\mathtt{prf}_{\boldsymbol{\tau}}(p, \ulcorner \langle\langle \mathcal{K}(\overline{u}) > \overline{\mathfrak{c}_{\boldsymbol{\tau}}} \rangle\rangle \urcorner)$ holds. Then, $\mathcal{T}\!\mathit{h}_{\boldsymbol{\tau}} \vdash \langle\langle \mathcal{K}(\overline{u}) > \overline{\mathfrak{c}_{\boldsymbol{\tau}}} \rangle\rangle$ and also $\boldsymbol{\varphi}_{\mathfrak{c}_{\boldsymbol{\tau}}}(0) = u$ holds; thus, $\mathcal{K}(u) \leqslant \mathfrak{c}_{\boldsymbol{\tau}}$. Whence, the Σ_1-sentence $\langle\langle \mathcal{K}(\overline{u}) \leqslant \overline{\mathfrak{c}_{\boldsymbol{\tau}}} \rangle\rangle$ is true, and so should be provable in $\mathcal{T}\!\mathit{h}_{\boldsymbol{\tau}}$; a contradiction with the consistency of $\mathcal{T}\!\mathit{h}_{\boldsymbol{\tau}}$ (cf. [25, Theorem 3.3]). Let us note that there are cofinitely many w's such that $\mathbb{N} \vDash \langle\langle \mathcal{K}(\overline{w}) > \overline{\mathfrak{c}_{\boldsymbol{\tau}}} \rangle\rangle$; so it is tempting to fix one of those w's as $\mathfrak{w}_{\boldsymbol{\tau}}$, and define *the Chaitin sentence of system* $\boldsymbol{\tau}$ as

$\mathbb{C}_\tau = \langle\langle \mathcal{K}(\overline{\mathfrak{w}_\tau}) > \overline{\mathfrak{c}_\tau}\rangle\rangle$. For technical reasons, we will postpone this till the next subsection (Definition 3.12); for the moment, we would like our results on Chaitin's proof to be as general as possible. We now show a Rosserian form of Chaitin's theorem (see [25, Theorem 3.9]).

THEOREM 3.2. *For every consistent system $\boldsymbol{\tau}$ and for every $e \geqslant \mathfrak{c}_\tau$, there are cofinitely many w's such that $\langle\langle \mathcal{K}(\overline{w}) > \overline{e}\rangle\rangle$ is independent from $\mathcal{TH}_\tau$.*

PROOF. Fix an $e \geqslant \mathfrak{c}_\tau$; we showed that for no w can $\mathcal{TH}_\tau \vdash \langle\langle \mathcal{K}(\overline{w}) > \overline{e}\rangle\rangle$ hold. We now show that $\mathcal{TH}_\tau \vdash \langle\langle \mathcal{K}(\overline{w}) \leqslant \overline{e}\rangle\rangle$ can hold for at most $(e+1)$-many w's. First let us note that the Δ_0-sentence $\forall\{x_i < \overline{n}\}_{i\leqslant n}\big(\bigvee_{i<j\leqslant n} x_i = x_j\big)$, which is a version of the Pigeonhole Principle, is true for each $n \in \mathbb{N}$, and thus is provable in $\mathfrak{B}$. Reason inside $\mathcal{TH}_\tau$:

> If for some distinct $\overline{\mathsf{w}_0}, \overline{\mathsf{w}_1}, \ldots, \overline{\mathsf{w}_{e+1}}$ we have $\langle\langle \mathcal{K}(\overline{\mathsf{w}_i}) \leqslant \overline{e}\rangle\rangle$, then for each $i \leqslant e+1$ we have $\langle\langle \varphi_{\overline{z_i}}(0) = \overline{\mathsf{w}_i}\rangle\rangle$ for some $z_i \leqslant e$. By the Pigeonhole Principle there should exist some $j < k \leqslant e+1$ such that $\overline{z_j} = \overline{z_k}$. Thus, we should have $\overline{\mathsf{w}_j} = \overline{\mathsf{w}_k}$, contradicting the distinctness of $\overline{\mathsf{w}_i}$'s.

So, there are cofinitely many w's for which we have both $\mathcal{TH}_\tau \nvdash \langle\langle \mathcal{K}(\overline{w}) > \overline{e}\rangle\rangle$ and $\mathcal{TH}_\tau \nvdash \langle\langle \mathcal{K}(\overline{w}) \leqslant \overline{e}\rangle\rangle$. ⊣

We note that if $\mathcal{TH}_\tau \nvdash \langle\langle \mathcal{K}(\overline{w}) \leqslant \overline{e}\rangle\rangle$, then $\mathbb{N} \vDash \langle\langle \mathcal{K}(\overline{w}) > \overline{e}\rangle\rangle$. So, it would be more tempting to fix as $\mathfrak{w}_\tau$ one of the cofinitely many w's with $\mathcal{TH}_\tau \nvdash \langle\langle \mathcal{K}(\overline{w}) \leqslant \overline{e}\rangle\rangle$; since then the Chaitin sentence $\mathbb{C}_\tau$ of $\boldsymbol{\tau}$ will be independent from $\mathcal{TH}_\tau$. In the following theorem, we show that Chaitin's theorem can deliver no constructive Π_1-incompleteness witness. Let us note that adding a true Π_1-sentence to a Σ_1-sound theory results in a Σ_1-sound theory; and the union of a chain of Σ_1-sound theories is also a Σ_1-sound theory.

THEOREM 3.3. *For every constructive Π_1-incompleteness witness $\mathcal{F}$, and for every Σ_1-sound system $\boldsymbol{\tau}$ there exists some Σ_1-sound super-system $\boldsymbol{\varrho}$ of $\boldsymbol{\tau}$ such that we have $\mathcal{TH}_\varrho \nvdash \mathcal{F}_\varrho \to \langle\langle \mathcal{K}(\overline{w}) > \overline{\mathfrak{c}_\varrho}\rangle\rangle$ for all w.*

PROOF. Let $\boldsymbol{\tau}^0 = \boldsymbol{\tau}$, and inductively, $\boldsymbol{\tau}^{n+1}(x) = \boldsymbol{\tau}^n(x) \vee [x = \mathcal{F}_{\boldsymbol{\tau}^n}]$. We note that all the $\{\mathcal{TH}_{\boldsymbol{\tau}^n}\}_n$'s and also $\mathcal{TH}_{\boldsymbol{\tau}^\infty} = \bigcup_n \mathcal{TH}_{\boldsymbol{\tau}^n}$ are Σ_1-sound. We show that there is some n such that $\mathcal{TH}_{\boldsymbol{\tau}^n} \nvdash \mathcal{F}_{\boldsymbol{\tau}^n} \to \langle\langle \mathcal{K}(\overline{w}) > \overline{\mathfrak{c}_{\boldsymbol{\tau}^n}}\rangle\rangle$ for all w; thus proving the theorem. Assume, for the sake of a contradiction, that for every n there exists some w_n such that $\mathcal{TH}_{\boldsymbol{\tau}^n} \vdash \mathcal{F}_{\boldsymbol{\tau}^n} \to \langle\langle \mathcal{K}(\overline{w_n}) > \overline{\mathfrak{c}_{\boldsymbol{\tau}^n}}\rangle\rangle$. There are two cases:

(1) For some n, $\mathfrak{c}_{\boldsymbol{\tau}^n} \geqslant \mathfrak{c}_{\boldsymbol{\tau}^{n+1}}$. Then, from $\mathcal{TH}_{\boldsymbol{\tau}^{n+1}} \vdash \langle\langle \mathcal{K}(\overline{w_n}) > \overline{\mathfrak{c}_{\boldsymbol{\tau}^n}}\rangle\rangle$ we have $\mathcal{TH}_{\boldsymbol{\tau}^{n+1}} \vdash \langle\langle \mathcal{K}(\overline{w_n}) > \overline{\mathfrak{c}_{\boldsymbol{\tau}^{n+1}}}\rangle\rangle$, contradicting Chaitin's theorem for $\boldsymbol{\tau}^{n+1}$.
(2) For every n, $\mathfrak{c}_{\boldsymbol{\tau}^n} < \mathfrak{c}_{\boldsymbol{\tau}^{n+1}}$. By the constructivity of $\mathcal{F}$ the system $\mathcal{TH}_{\boldsymbol{\tau}^\infty}$ is RE. So, there should exist some m such that $\mathfrak{c}_{\boldsymbol{\tau}^\infty} \leqslant \mathfrak{c}_{\boldsymbol{\tau}^m}$. Thus, from $\mathcal{TH}_{\boldsymbol{\tau}^{m+1}} \vdash \langle\langle \mathcal{K}(\overline{w_m}) > \overline{\mathfrak{c}_{\boldsymbol{\tau}^m}}\rangle\rangle$, we get $\mathcal{TH}_{\boldsymbol{\tau}^\infty} \vdash \langle\langle \mathcal{K}(\overline{w_m}) > \overline{\mathfrak{c}_{\boldsymbol{\tau}_\infty}}\rangle\rangle$, which contradicts Chaitin's theorem for the consistent $\boldsymbol{\tau}^\infty$. ⊣

A mapping $\boldsymbol{\tau} \mapsto w_\tau$, which assigns $w_\tau \in \mathbb{N}$ to a given system $\boldsymbol{\tau}$, is called a *Chaitin mapping* when $\mathbb{N} \vDash \langle\langle \mathcal{K}(\overline{w_\tau}) > \overline{\mathfrak{c}_\tau}\rangle\rangle$ holds for every consistent system $\boldsymbol{\tau}$. Let us call it a *Rosser–Chaitin* mapping when $\mathcal{TH}_\tau \nvdash \langle\langle \mathcal{K}(\overline{w_\tau}) \leqslant \overline{\mathfrak{c}_\tau}\rangle\rangle$ holds for every consistent system $\boldsymbol{\tau}$.

COROLLARY 3.4. *No Chaitin mapping can be constructive* ([25, *Theorem* 3.5]).

Chaitin's Π_1*-incompleteness theorem can deliver none of the* Π_1*-incompleteness theorems of Gödel* (*first and second*), *Kleene* (*first and second*), *or Rosser.* ⊣

We will see below that Chaitin's Π_1-incompleteness witness does deliver a variant of the Π_1-incompleteness witness of Boolos; thus, Boolos' theorem is not constructive (see [25, Theorem 4.5]). As the last result on Chaitin's theorem, we show that essentially no Π_1-incompleteness witness may deliver Chaitin's Π_1-incompleteness witness in the standard way:

COROLLARY 3.5. *For every consistent system* $\boldsymbol{\tau}$ *and every sentence* ψ *with* $\mathcal{T}\!\mathit{h}_{\boldsymbol{\tau}} \nvdash \psi$, *there are cofinitely many* w*'s with* $\mathcal{T}\!\mathit{h}_{\boldsymbol{\tau}} \nvdash \langle\langle \mathcal{K}(\overline{w}) > \overline{\mathfrak{c}_{\boldsymbol{\tau}}}\rangle\rangle \to \psi$.

PROOF. Since $\mathcal{T}\!\mathit{h}_{\boldsymbol{\tau}} + \neg\psi$ is consistent, then by (the proof of) Theorem 3.2 for cofinitely many w's, we have $\mathcal{T}\!\mathit{h}_{\boldsymbol{\tau}} + \neg\psi \nvdash \langle\langle \mathcal{K}(\overline{w}) \leqslant \overline{\mathfrak{c}_{\boldsymbol{\tau}}}\rangle\rangle$. ⊣

As a consequence, for every given Rosser–Chaitin mapping $\boldsymbol{\tau} \mapsto v_{\boldsymbol{\tau}}$, there exists another Rosser–Chaitin mapping $\boldsymbol{\tau} \mapsto w_{\boldsymbol{\tau}}$ such that for every consistent system $\boldsymbol{\tau}$ we have $\mathcal{T}\!\mathit{h}_{\boldsymbol{\tau}} \nvdash \langle\langle \mathcal{K}(\overline{w_{\boldsymbol{\tau}}}) > \overline{\mathfrak{c}_{\boldsymbol{\tau}}}\rangle\rangle \to \langle\langle \mathcal{K}(\overline{v_{\boldsymbol{\tau}}}) > \overline{\mathfrak{c}_{\boldsymbol{\tau}}}\rangle\rangle$.

3.2. Boolos' proof. Finally, we consider the theorem of Boolos, for which we make the following convention.

CONVENTION. *All the variables are* $\boldsymbol{\vartheta}, \boldsymbol{\vartheta}', \boldsymbol{\vartheta}'', \boldsymbol{\vartheta}''', \ldots$ *whose lengths are* $1, 2, 3, 4, \ldots$, *respectively.* ⊣

The length of an expression is the number of symbols in it. By the above convention (3.6), for any natural number $n \in \mathbb{N}$, there are at most finitely many formulas with length n; without this convention, for variables $x, y, z, \ldots$, all of the formulas $x = 0$, $y = 0$, $z = 0, \ldots$ would be length of 3.

DEFINITION 3.7. A number $n \in \mathbb{N}$ is *definable* in the theory T, by the formula $\theta(\boldsymbol{\vartheta})$ with the only free variable $\boldsymbol{\vartheta}$, when $T \vdash \forall\boldsymbol{\vartheta}[\theta(\boldsymbol{\vartheta}) \leftrightarrow \boldsymbol{\vartheta} = \bar{n}]$ holds. Let $\boldsymbol{\delta}(\ulcorner\theta\urcorner, n)$ denote the Gödel code $\ulcorner\forall\boldsymbol{\vartheta}[\theta(\boldsymbol{\vartheta}) \leftrightarrow \boldsymbol{\vartheta} = \bar{n}]\urcorner$. ⊣

Suppose that the formula $\mathtt{F}_1(x)$ in the language of arithmetic states that "x is (the Gödel code of) a formula which has $\boldsymbol{\vartheta}$ as its only free variable," and the formula $\mathtt{L}^{<y}(x)$ states that "the formula (with Gödel code) x has length less than y." Indeed, there are such Σ_1-formulas in the language of arithmetic, whose existence can be shown by the techniques of Gödel's arithmetization.

DEFINITION 3.8. For a system $\boldsymbol{\tau}$, let $\mathcal{D}_{\boldsymbol{\tau}}^{<y}(x) = \exists\xi[\mathtt{F}_1(\xi) \wedge \mathtt{L}^{<y}(\xi) \wedge \mathtt{Pr}_{\boldsymbol{\tau}}(\boldsymbol{\delta}(\xi, x))]$ be a formula, in the language of arithmetic, stating that "x is definable in theory $\mathcal{T}\!\mathit{h}_{\boldsymbol{\tau}}$ by a formula with length $< y$." Let $\mathcal{B}_{\boldsymbol{\tau}}^{<y}(x) = \neg\mathcal{D}_{\boldsymbol{\tau}}^{<y}(x) \wedge \forall z < x\, \mathcal{D}_{\boldsymbol{\tau}}^{<y}(z)$ be the formula which states that "x is the least number not definable (in $\mathcal{T}\!\mathit{h}_{\boldsymbol{\tau}}$) by any formula with length less than y." Let $\boldsymbol{\ell}_{\boldsymbol{\tau}}$ be the length of $\mathcal{B}_{\boldsymbol{\tau}}^{<\vartheta'}(\boldsymbol{\vartheta})$. ⊣

Let k be any natural number nonsmaller than 10 (Boolos [5] originally takes it to be 10). For a system $\boldsymbol{\tau}$, let $\mathbb{k}_{\boldsymbol{\tau}} = \bar{k} \cdot \bar{\boldsymbol{\ell}}_{\boldsymbol{\tau}}$ be a term representing $k\boldsymbol{\ell}_{\boldsymbol{\tau}}$. Let $\boldsymbol{\beta}_{\boldsymbol{\tau}}(\vartheta) = \exists \vartheta' [\vartheta' = \mathbb{k}_{\boldsymbol{\tau}} \wedge \mathcal{B}_{\boldsymbol{\tau}}^{<\vartheta'}(\vartheta)]$ be the formula stating that "ϑ is the least number that is not definable by any formula with length less than $k\boldsymbol{\ell}_{\boldsymbol{\tau}}$." It can be shown that the length of $\boldsymbol{\beta}_{\boldsymbol{\tau}}$ is less than $k\boldsymbol{\ell}_{\boldsymbol{\tau}}$ (cf. the proof of Theorem 4.3 in [25]). Let $\mathfrak{b}_{\boldsymbol{\tau}}$ be the least number (if any) that is not definable (in $\mathcal{T}\hbar_{\boldsymbol{\tau}}$) by any formula with length less than $k\boldsymbol{\ell}_{\boldsymbol{\tau}}$. Boolos' original theorem is the following:

THEOREM 3.9. *If $\boldsymbol{\tau}$ is a consistent system, then $\boldsymbol{\beta}_{\boldsymbol{\tau}}(\overline{\mathfrak{b}_{\boldsymbol{\tau}}})$ is a true sentence that is not provable in $\mathcal{T}\hbar_{\boldsymbol{\tau}}$.*

PROOF. The truth of $\boldsymbol{\beta}_{\boldsymbol{\tau}}(\overline{\mathfrak{b}_{\boldsymbol{\tau}}})$ follows from the definition of $\mathfrak{b}_{\boldsymbol{\tau}}$. Assume, for the sake of a contradiction, that $\mathcal{T}\hbar_{\boldsymbol{\tau}} \vdash \boldsymbol{\beta}_{\boldsymbol{\tau}}(\overline{\mathfrak{b}_{\boldsymbol{\tau}}})$; thus, $\mathcal{T}\hbar_{\boldsymbol{\tau}} \vdash \mathcal{B}_{\boldsymbol{\tau}}^{<\mathbb{k}_{\boldsymbol{\tau}}}(\overline{\mathfrak{b}_{\boldsymbol{\tau}}})$. Then, $\mathfrak{B} \vdash \forall x [\mathcal{B}_{\boldsymbol{\tau}}^{<\mathbb{k}_{\boldsymbol{\tau}}}(x) \leftrightarrow x = \overline{\mathfrak{b}_{\boldsymbol{\tau}}}]$,[3] and so $\mathfrak{b}_{\boldsymbol{\tau}}$ is definable (in $\mathcal{T}\hbar_{\boldsymbol{\tau}}$) by the formula $\mathcal{B}_{\boldsymbol{\tau}}^{<\mathbb{k}_{\boldsymbol{\tau}}}(\vartheta)$ whose length is less than $k\boldsymbol{\ell}_{\boldsymbol{\tau}}$; thus, $\mathcal{T}\hbar_{\boldsymbol{\tau}}$ is inconsistent. ⊣

Let us note that $\boldsymbol{\beta}_{\boldsymbol{\tau}}(\overline{\mathfrak{b}_{\boldsymbol{\tau}}})$ is not a Π_1-sentence; though we have

$$\mathfrak{B} \vdash \boldsymbol{\beta}_{\boldsymbol{\tau}}(\overline{\mathfrak{b}_{\boldsymbol{\tau}}}) \equiv \mathcal{B}_{\boldsymbol{\tau}}^{<\mathbb{k}_{\boldsymbol{\tau}}}(\overline{\mathfrak{b}_{\boldsymbol{\tau}}}) \equiv \neg \mathcal{D}_{\boldsymbol{\tau}}^{<\mathbb{k}_{\boldsymbol{\tau}}}(\overline{\mathfrak{b}_{\boldsymbol{\tau}}}) \wedge \forall z < \overline{\mathfrak{b}_{\boldsymbol{\tau}}}\, \mathcal{D}_{\boldsymbol{\tau}}^{<\mathbb{k}_{\boldsymbol{\tau}}}(z) \equiv \neg \mathcal{D}_{\boldsymbol{\tau}}^{<\mathbb{k}_{\boldsymbol{\tau}}}(\overline{\mathfrak{b}_{\boldsymbol{\tau}}}),$$

because $\forall z < \overline{\mathfrak{b}_{\boldsymbol{\tau}}}\, \mathcal{D}_{\boldsymbol{\tau}}^{<\mathbb{k}_{\boldsymbol{\tau}}}(z)$ is a true Σ_1-sentence, for a consistent system $\boldsymbol{\tau}$, and so it is $\mathfrak{B}$-provable. Whence, the essence of Boolos' theorem is the truth and $\mathcal{T}\hbar_{\boldsymbol{\tau}}$-unprovability of the Π_1-sentence $\neg \mathcal{D}_{\boldsymbol{\tau}}^{<\mathbb{k}_{\boldsymbol{\tau}}}(\overline{\mathfrak{b}_{\boldsymbol{\tau}}})$. Indeed, $\mathfrak{G}_{\mathbb{2}}$ can deliver this, and much more:

THEOREM 3.10. *For every consistent system $\boldsymbol{\tau}$ and every numbers $m,n \in \mathbb{N}$ with $m > 3$ we have $\mathcal{T}\hbar_{\boldsymbol{\tau}} \nvdash \neg \mathcal{D}_{\boldsymbol{\tau}}^{<\overline{m}}(\overline{n})$.*

PROOF. If $\mathcal{T}\hbar_{\boldsymbol{\tau}} \vdash \neg \mathcal{D}_{\boldsymbol{\tau}}^{<\overline{m}}(\overline{n})$, then $\mathcal{T}\hbar_{\boldsymbol{\tau}} \vdash \forall \xi [\mathtt{F}_1(\xi) \wedge \mathtt{L}^{<\overline{m}}(\xi) \rightarrow \neg \mathtt{Pr}_{\boldsymbol{\tau}}(\boldsymbol{\delta}(\xi, \overline{n}))]$. There is some formula $\zeta(\vartheta)$ with the only free variable ϑ, such as $\vartheta = 0$, whose length is less than m; thus, $\mathcal{T}\hbar_{\boldsymbol{\tau}} \vdash \neg \mathtt{Pr}_{\boldsymbol{\tau}}(\boldsymbol{\delta}(\ulcorner \zeta \urcorner, \overline{n}))$. So, by Lemma 2.1(1), we should have $\mathcal{T}\hbar_{\boldsymbol{\tau}} \vdash \mathtt{Con}_{\boldsymbol{\tau}}$, which contradicts $\mathfrak{G}_{\mathbb{2}}$. ⊣

Chaitin's theorem can deliver Boolos' theorem too, even in a more general form:

THEOREM 3.11. *For every consistent system $\boldsymbol{\tau}$ there is a number $\mathfrak{k}_{\boldsymbol{\tau}}$ such that for every system $\boldsymbol{\varrho}$ and for every m,n with $m \geqslant \mathfrak{k}_{\boldsymbol{\tau}}$ we have $\mathcal{T}\hbar_{\boldsymbol{\tau}} \nvdash \neg \mathcal{D}_{\boldsymbol{\varrho}}^{<\overline{m}}(\overline{n})$.*

PROOF. Let $\mathfrak{k}_{\boldsymbol{\tau}}$ be a number that is greater than the lengths of all the formulas $\langle\langle \boldsymbol{\varphi}_{\overline{e}}(0) = \vartheta \rangle\rangle$ with $e \leqslant \mathfrak{c}_{\boldsymbol{\tau}}$. Fix $m,n \in \mathbb{N}$ with $m \geqslant \mathfrak{k}_{\boldsymbol{\tau}}$. We first show (even inside $\mathfrak{B}$) that (for an arbitrary system $\boldsymbol{\varrho}$) if $\neg \mathcal{D}_{\boldsymbol{\varrho}}^{<m}(n)$, then $\mathcal{K}(n) > \mathfrak{c}_{\boldsymbol{\tau}}$: If, on the contrary, we had $\mathcal{K}(n) \leqslant \mathfrak{c}_{\boldsymbol{\tau}}$, then for some $e \leqslant \mathfrak{c}_{\boldsymbol{\tau}}$ we would have $\boldsymbol{\varphi}_e(0) = n$. Since $\mathfrak{B}$ can strongly represent all the recursive functions, then n would be definable in $\mathfrak{B}$ (and so in $\mathcal{T}\hbar_{\boldsymbol{\varrho}}$) by the formula $\langle\langle \boldsymbol{\varphi}_{\overline{e}}(0) = \vartheta \rangle\rangle$ whose length is less than $\mathfrak{k}_{\boldsymbol{\tau}} \leqslant m$. Thus, $\mathcal{D}_{\boldsymbol{\varrho}}^{<\mathfrak{k}_{\boldsymbol{\tau}}}(n)$ and so $\mathcal{D}_{\boldsymbol{\varrho}}^{<m}(n)$ would hold;

[3] Here, we use the fact that $\mathfrak{B} \vdash \forall x (x < \overline{n} \vee x = \overline{n} \vee \overline{n} < x)$ for each $n \in \mathbb{N}$ (see the proof of Theorem 4.3 in [25]).

a contradiction. Therefore, $\mathcal{Th}_{\tau} \nvdash \neg \mathcal{D}_{\varrho}^{<\overline{m}}(\overline{n})$ follows from Chaitin's theorem that $\mathcal{Th}_{\tau} \nvdash \langle\langle \mathcal{K}(\overline{n}) > \overline{\mathfrak{c}_{\tau}} \rangle\rangle$. ⊣

Now we can define the Chaitin and the Boolos sentence(s) of a system.

Definition 3.12. For a system $\boldsymbol{\tau}$, if $\mathcal{Th}_{\tau}$ is not consistent, then let $\mathfrak{w}_{\tau} = 0$. If $\mathcal{Th}_{\tau}$ is consistent, then let $\mathfrak{w}_{\tau}$ be one of the cofinitely many w's such that

(i) $\mathcal{Th}_{\tau} + \neg \overline{\mathbb{R}}_{\tau} \nvdash \langle\langle \mathcal{K}(\overline{w}) \leqslant \overline{\mathfrak{c}_{\tau}} \rangle\rangle$, and
(ii) $\mathbb{N} \vDash \neg \mathcal{D}_{\tau}^{<\overline{\kappa_{\tau}}}(\overline{w})$;

where, $\boldsymbol{\kappa}_{\tau}$ is the least number k such that $k \geqslant 10 \cdot \boldsymbol{\ell}_{\tau}$, and also k is greater than the lengths of all the formulas $\langle\langle \boldsymbol{\varphi}_{\overline{e}}(0) = \boldsymbol{\vartheta} \rangle\rangle$ with $\mathrm{e} \leqslant \mathfrak{c}_{\tau}$.
For a consistent system $\boldsymbol{\tau}$, let $\mathbb{C}_{\tau} = \langle\langle \mathcal{K}(\overline{\mathfrak{w}_{\tau}}) > \overline{\mathfrak{c}_{\tau}} \rangle\rangle$ be *the Chaitin sentence* of $\boldsymbol{\tau}$, and let $\widetilde{\mathbb{B}}_{\tau} = \neg \mathcal{D}_{\tau}^{<\overline{\kappa_{\tau}}}(\overline{\mathfrak{w}_{\tau}})$ be (a variant of) *the Boolos sentence* of $\boldsymbol{\tau}$. ⊣

Let us note that for every consistent system $\boldsymbol{\tau}$, the Π_1-sentences $\mathbb{C}_{\tau}$ and $\widetilde{\mathbb{B}}_{\tau}$ are both true and unprovable in $\mathcal{Th}_{\tau}$.

Corollary 3.13. (1) *For every consistent system* $\boldsymbol{\tau}$, $\mathcal{Th}_{\tau} \vdash \widetilde{\mathbb{B}}_{\tau} \to \mathbb{C}_{\tau}$;
(2) *For every consistent system* $\boldsymbol{\tau}$, $\mathcal{Th}_{\tau} \vdash \widetilde{\mathbb{B}}_{\tau} \to \mathtt{Con}_{\tau}$;
(3) *Boolos'* Π_1*-incompleteness theorem is not constructive.*

Proof. (1) The deduction $\mathfrak{B} \vdash \neg \mathcal{D}_{\tau}^{<\overline{\kappa_{\tau}}}(\overline{\mathfrak{w}_{\tau}}) \to \langle\langle \mathcal{K}(\overline{\mathfrak{w}_{\tau}}) > \overline{\mathfrak{c}_{\tau}} \rangle\rangle$ was shown in the proof of Theorem 3.11. (2) Follows from the proof of Theorem 3.10. (3) Follows from Theorem 3.3 and the item (1) above. ⊣

Corollary 3.14. *For every consistent system* $\boldsymbol{\tau}$,

(1) $\mathcal{Th}_{\tau} \nvdash \mathbb{C}_{\tau} \to \overline{\mathbb{R}}_{\tau}$;
(2) $\mathcal{Th}_{\tau} \nvdash \mathbb{C}_{\tau} \to \mathtt{Con}_{\tau}$;
(3) $\mathcal{Th}_{\tau} \nvdash \mathbb{C}_{\tau} \to \widetilde{\mathbb{B}}_{\tau}$;
(4) $\mathcal{Th}_{\tau} \nvdash \overline{\mathbb{R}}_{\tau} \to \widetilde{\mathbb{B}}_{\tau}$.

Proof. (1) By Definition 3.12. (2) By Theorem 2.4 and (1) above. (3) By Corollary 3.13(2) and (2) above. (4) By Theorem 2.4 and Corollary 3.13(2). ⊣

Corollary 3.15. *For every* Σ_1*-sound system* $\boldsymbol{\tau}$ *there exists some* Σ_1*-sound super-system* $\boldsymbol{\varrho}$ *of* $\boldsymbol{\tau}$ *such that*

(1) $\mathcal{Th}_{\varrho} \nvdash \mathtt{Con}_{\varrho} \to \mathbb{C}_{\varrho}$;
(2) $\mathcal{Th}_{\varrho} \nvdash \overline{\mathbb{R}}_{\varrho} \to \mathbb{C}_{\varrho}$;
(3) $\mathcal{Th}_{\varrho} \nvdash \mathtt{Con}_{\varrho} \to \widetilde{\mathbb{B}}_{\varrho}$.

Proof. (1) By Theorem 3.3. (2) By Theorem 2.4 and (1) above. (3) By Corollary 3.13(1) and (1) above. ⊣

§4. Conclusions. We examined the incompleteness theorems of five great minds of symbolic logic, namely Gödel, Rosser, Kleene, Chaitin, and Boolos. We compared their proofs with each other, putting Gödel's second incompleteness theorem at the center of our attention, which resulted in the

following diagram (where $\mathcal{F} \preccurlyeq \mathcal{H}$ means that $\mathcal{F}$ is derived from $\mathcal{H}$ in the standard way and thus $\mathcal{F} \not\preccurlyeq \mathcal{H}$ means that $\mathcal{F}$ is not derivable from $\mathcal{H}$ in the standard way):

$$\mathbb{G}\mathtt{ödel}_{\mathbb{2}} \preccurlyeq \mathbb{G}\mathtt{ödel}_{\mathbb{1}} \preccurlyeq \mathbb{K}\mathtt{leene} \preccurlyeq \mathbb{G}\mathtt{ödel}_{\mathbb{2}} \qquad \overline{\mathbb{R}}\mathtt{osser} \preccurlyeq \overline{\mathbb{K}'}\mathtt{leene} \preccurlyeq \overline{\mathbb{R}}\mathtt{osser}$$

$$\mathbb{G}\mathtt{ödel}_{\mathbb{2}} \preccurlyeq \overline{\mathbb{R}}\mathtt{osser} \not\preccurlyeq \mathbb{G}\mathtt{ödel}_{\mathbb{2}} \qquad \overline{\mathbb{R}}\mathtt{osser} \not\preccurlyeq \mathbb{C}\mathtt{haitin} \not\preccurlyeq \overline{\mathbb{R}}\mathtt{osser}$$

$$\mathbb{G}\mathtt{ödel}_{\mathbb{2}} \not\preccurlyeq \mathbb{C}\mathtt{haitin} \not\preccurlyeq \mathbb{G}\mathtt{ödel}_{\mathbb{2}} \qquad \mathbb{G}\mathtt{ödel}_{\mathbb{2}} \not\preccurlyeq \widetilde{\mathbb{B}}\mathtt{oolos} \preccurlyeq \mathbb{G}\mathtt{ödel}_{\mathbb{2}}$$

$$\widetilde{\mathbb{B}}\mathtt{oolos} \preccurlyeq \mathbb{C}\mathtt{haitin} \not\preccurlyeq \widetilde{\mathbb{B}}\mathtt{oolos} \qquad \widetilde{\mathbb{B}}\mathtt{oolos} \preccurlyeq \overline{\mathbb{R}}\mathtt{osser} \not\preccurlyeq \widetilde{\mathbb{B}}\mathtt{oolos}$$

The lines over $\mathbb{R}$ and $\mathbb{K}'$ indicate that some alternative versions of the sentences of Rosser and (the second) Kleene have been considered, and the tilde over $\mathbb{B}$ indicates that a substantial variant of the sentence of Boolos is considered.

As the diagram shows, Boolos' theorem is indeed the weakest among the other theorems, since it is derivable from all of them. Rosser's theorem is the strongest in a sense, since it delivers all of the other theorems except Chaitin's incompleteness theorem. Chaitin's is the most neutral one, since it is not derived from any other theorem, and it delivers no other theorem, except Boolos'. Here, we did not study the incompleteness proofs whose unprovable sentences are not Π_1; one prominent example is Kripke's proof [21] for the incompleteness theorem, which shows the Π_2-incompleteness of Σ_2-sound and (RE) extensions of $\mathfrak{B}$.

Let us examine Boolos' original proof more closely from [5] to see the little and amusing point that Boolos' theorem is derivable from $\mathbb{G}_{\mathbb{2}}$: His formula $B(x,y)$ is our $\mathcal{D}_{\boldsymbol{\tau}}^{<y}(x)$, stating that "$x$ is definable (namable) by a formula with length $< y$ in theory $\mathcal{T}\hbar_{\boldsymbol{\tau}}$." His $A(x,y)$ is our $\mathcal{B}_{\boldsymbol{\tau}}^{<y}(x)$, stating that "$x$ is the least number not definable (not named) by any formula with length $< y$ in $\mathcal{T}\hbar_{\boldsymbol{\tau}}$." Boolos' k is our $\boldsymbol{\ell}_{\boldsymbol{\tau}}$, the length of $B(x,y)$, and his $F(x) = \exists y\big(y = [10]\times[k] \wedge A(x,y)\big)$ is our $\boldsymbol{\beta}_{\boldsymbol{\tau}}(x)$. Boolos notes that the length of $F(x)$ is less than $10k$, and if n (our $\mathfrak{b}_{\boldsymbol{\tau}}$) is the least number not definable by a formula with length less than $10k$, then $\forall x(F(x) \leftrightarrow x = [n])$ is true but not in the output of M (unprovable in our $\mathcal{T}\hbar_{\boldsymbol{\tau}}$). This sentence is not Π_1, but it is equivalent with $F([n])$, and this is equivalent with $A([n],[10]\times[k])$. This is not Π_1 either, but it is equivalent in M (or our $\mathcal{T}\hbar_{\boldsymbol{\tau}}$) with $\neg B([n],[10]\times[k])$, which is a Π_1-sentence. This sentence says that n is not definable by any formula with length less than $10k$, and in particular it is not definable by the formula $F(x)$. Thus, $\neg B([n],[10]\times[k])$ implies the unprovability of $F([n])$ in $\mathcal{T}\hbar_{\boldsymbol{\tau}}$ (that $F([n])$ is not in the output list of M), so it implies $\mathtt{Con}_{\boldsymbol{\tau}}$, the consistency of $\mathcal{T}\hbar_{\boldsymbol{\tau}}$ (that M does not output contradictory statements). Whence, $\forall x(F(x) \leftrightarrow x = [n])$ is not provable in $\mathcal{T}\hbar_{\boldsymbol{\tau}}$, because $\mathtt{Con}_{\boldsymbol{\tau}}$ is not $\mathcal{T}\hbar_{\boldsymbol{\tau}}$-provable by $\mathbb{G}_{\mathbb{2}}$. So, the unprovability of the Boolos sentence follows from $\mathbb{G}_{\mathbb{2}}$ (Boolos continues his argument in [5] and shows the unprovability of $F([n])$ by an argument similar to Berry's paradox; see the proof of Theorem 3.9).

One could read in the literature that $\mathbb{G}_{\mathbb{2}}$ follows from the first incompleteness theorem; this is said (and is true) for Gödel's proof, and we showed that it is true also for Rosser's proof and Kleene's proof(s). As the history goes,

the ground breaking paper [9] of Gödel was the first part, as its title shows. The second part never appeared, as Gödel felt that people could derive $\mathbb{G}_2$ (which was promised to be proven in a sequel paper) by themselves from the first theorem; so he did not even attempt to write it. On some other proofs for the first incompleteness theorem, one may read the opposite; for example, the authors of [14] write that Maehara [20] "insists that Boolos' theorem is different from Gödel's one," one reason being that "we cannot obtain the second theorem from Boolos' theorem in the standard way." We gave a rigorous proof for this insight in Corollary 3.15(3), and showed, moreover, that one cannot obtain $\mathbb{G}_2$ from Chatin's theorem, in the standard way, either.

Acknowledgments. This research was partially supported by the grant No. 96030030 of the Institute for Research in Fundamental Sciences (IPM), Tehran, Iran. The most helpful comments and suggestions of the referee remarkably improved the presentation and the strength of the results; this is highly appreciated.

REFERENCES

[1] S. Artemov and L. Beklemishev, *Provability logic*, ***Handbook of Philosophical Logic*** (D. Gabbay and F. Guenthner, editors), second ed., Springer, New York, 2005, pp. 189–360.

[2] A. Avron, *A note of provability, truth and existence*. ***Journal of Philosophical Logic***, vol. 20 (1991), no. 4, pp. 403–409.

[3] L. Beklemishev, *Reflection principles and provability algebras in formal arithmetic*. ***Russian Mathematical Surveys***, vol. 60 (2005), no. 2, pp. 197–268.

[4] R. Blanck, ***On Rosser sentences and proof predicates***, M.A. thesis, Department of Philosophy, University of Göteborg, Sweden, 2006. Available at http://bit.do/fxUqf.

[5] G. Boolos, *A new proof of the Gödel incompleteness theorem*. ***Notices of the American Mathematical Society***, vol. 36 (1989), no. 4, pp. 388–390. *A Letter from George Boolos*, ibid 36 (1989), no. 6, p. 676.

[6] ———, ***The Logic of Provability***, Cambridge University Press, Cambridge, 1994.

[7] G. Chaitin, *Computational complexity and Gödel's incompleteness theorem*. ***SIGACT News***, vol. 9 (1971), pp. 11–12. Abstract in ***Notices of the American Mathematical Society***, vol. 17 (1970), no. 6, p. 672.

[8] W. Craig, *On axiomatizability within a system*. ***The Journal of Symbolic Logic***, vol. 18 (1953), no. 1, pp. 30–32.

[9] K. Gödel, *Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme, I.*, ***Monatshefte für Mathematik und Physik***, vol. 38 (1931), no. 1, pp. 173–198 (in German). Translated into English as: *On formally undecidable propositions of Principia Mathematica and Related Systems, I*, ***Kurt Gödel, Collected Works, Volume I: Publications 1929–1936*** (S. Feferman, J. W. Dawson Jr., S. C. Kleene, G. H. Moore, R. M. Solovay, and J. van Heijenoort, editors), Oxford University Press, Oxford, 1986, pp. 135–152.

[10] D. Guaspari and R. Solovay, *Rosser sentences*. ***Annals of Mathematical Logic***, vol. 16 (1979), no. 1, pp. 81–99.

[11] P. Hájek and P. Pudlák, ***Metamathematics of First-Order Arithmetic***, second ed., Springer-Verlag, New York, 1998.

[12] D. Isaacson, *Necessary and sufficient conditions for undecidability of the Gödel sentence and its truth*, ***Logic, Mathematics, Philosophy: Vintage Enthusiasms—Essays in Honour of John L. Bell*** (D. DeVidi, M. Hallett, and P. Clarke, editors), Springer, New York, 2011, pp. 135–152.

[13] R. Kaye, ***Models of Peano Arithmetic***, Oxford University Press, Oxford, 1991.

[14] M. Kikuchi, T. Kurahashi, and H. Sakai, *On proofs of the incompleteness theorems based on Berry's paradox by Vopěnka, Chaitin, and Boolos*. ***Mathematical Logic Quarterly***, vol. 58 (2012), no. 4–5, pp. 307–316.

[15] S. Kleene, *General recursive functions of natural numbers*. ***Mathematische Annalen***, vol. 112 (1936), no. 1, pp. 727–742.

[16] ———, *A symmetric form of Gödel's theorem*. ***Indagationes Mathematicae***, vol. 12 (1950), pp. 244–246.

[17] G. Kreisel, *On weak completeness of intuitionistic predicate logic*. ***The Journal of Symbolic Logic***, vol. 27 (1962), no. 2, pp. 139–158.

[18] G. Kreisel and G. Takeuti, *Formally self-referential propositions for cut free classical analysis and related systems*. ***Dissertationes Mathematic***æ, vol. 118 (1974), pp. 1–50.

[19] A. Macintyre and H. Simmons, *Gödel's diagonalization technique and related properties of theories*. ***Colloquium Mathematicum***, vol. 28 (1973), pp. 165–180.

[20] S. Maehara, *Boolos Shi No Genkou Wo Mite*. ***Gendai Shisou*** (December 1989), pp. 80–92.

[21] H. Putnam, *Nonstandard models and Kripke's proof of the Gödel theorem*. ***Notre Dame Journal of Formal Logic***, vol. 41 (2000), no. 1, pp. 53–58.

[22] B. Rosser, *Extensions of some theorems of Gödel and Church*. ***The Journal of Symbolic Logic***, vol. 1 (1936), no. 3, pp. 87–91.

[23] S. Salehi, *Gödel's incompleteness phenomenon—Computationally*. ***Philosophia Scientiæ***, vol. 18 (2014), no. 3, pp. 22–37.

[24] S. Salehi and P. Seraji, *Gödel–Rosser's incompleteness theorem, generalized and optimized for definable theories*. ***Journal of Logic and Computation***, vol. 27 (2017), no. 5, pp. 1391–1397.

[25] ———, *On constructivity and the Rosser property: A closer look at some Gödelean proofs*. ***Annals of Pure and Applied Logic***, vol. 169 (2018), no. 10, pp. 971–980.

[26] T. A. Slaman, *Σ_n-bounding and Δ_n-induction*. ***Proceedings of The American Mathematical Society***, vol. 132 (2004), no. 8, pp. 2449–2456.

[27] C. Smoryński, ***Self-Reference and Modal Logic***, Springer, New York, 1985.

[28] A. Tarski, A. Mostowski, and R. M. Robinson, ***Undecidable Theories***, North-Holland, Netherlands, 1953. Reprinted by Dover Publications, 2010.

[29] A. Visser, *Another look at the second incompleteness theorem*. ***The Review of Symbolic Logic,*** vol. 13 (2020) no. 2, pp. 269–295.

[30] C. von Bülow, *A remark on equivalent Rosser sentences*. ***Annals of Pure and Applied Logic***, vol. 151 (2008), no. 1, pp. 62–67.

RESEARCH INSTITUTE FOR FUNDAMENTAL SCIENCES
UNIVERSITY OF TABRIZ
29 BAHMAN BOULEVARD, P.O. BOX 51666-17766, TABRIZ, IRAN
and
SCHOOL OF MATHEMATICS
INSTITUTE FOR RESEARCH IN FUNDAMENTAL SCIENCES
P.O. BOX 19395-5746, TEHRAN, IRAN
E-mail: root@saeedsalehi.ir
URL: http://saeedsalehi.ir/